ARTICLES
https://doi.org/10.1038/s41559-020-1102-0

nature ecology & evolution

There are amendments to this paper

The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean

Daniel M. Fernandes[1,2,3*], Alissa Mittnik[4], Iñigo Olalde[4], Iosif Lazaridis[4], Olivia Cheronet[1,2], Nadin Rohland[4], Swapan Mallick[4,5,6], Rebecca Bernardos[4], Nasreen Broomandkhoshbacht[4,5,32], Jens Carlsson[7], Brendan J. Culleton[8], Matthew Ferry[4,5], Beatriz Gamarra[2,9,10], Martina Lari[11], Matthew Mah[4,5,6], Megan Michel[4,5,33], Alessandra Modi[11], Mario Novak[2,12], Jonas Oppenheimer[4,5,34], Kendra A. Sirak[2,4,35], Kristin Stewardson[4,5], Kirsten Mandl[1], Constanze Schattke[1], Kadir T. Özdoğan[1], Michaela Lucci[13], Gabriella Gasperetti[14], Francesca Candilio[15], Gianfranca Salis[15], Stefania Vai[11], Edgard Camarós[16], Carla Calò[17], Giulio Catalano[18], Marián Cueto[16], Vincenza Forgia[19], Marina Lozano[9,10], Elisabetta Marini[17], Margherita Micheletti[20], Roberto M. Micciché[18], Maria R. Palombo[21], Damià Ramis[22], Vittoria Schimmenti[23], Pau Sureda[24,25], Luís Teira[16], Maria Teschler-Nicola[1,26], Douglas J. Kennett[27], Carles Lalueza-Fox[28], Nick Patterson[6,29], Luca Sineo[18], Alfredo Coppa[30], David Caramelli[11*], Ron Pinhasi[1,2*] and David Reich[4,5,6,29,31*]

Steppe-pastoralist-related ancestry reached Central Europe by at least 2500 BC, whereas Iranian farmer-related ancestry was present in Aegean Europe by at least 1900 BC. However, the spread of these ancestries into the western Mediterranean, where they have contributed to many populations that live today, remains poorly understood. Here, we generated genome-wide ancient-DNA data from the Balearic Islands, Sicily and Sardinia, increasing the number of individuals with reported data from 5 to 66. The oldest individual from the Balearic Islands (~2400 BC) carried ancestry from steppe pastoralists that probably derived from west-to-east migration from Iberia, although two later Balearic individuals had less ancestry from steppe pastoralists. In Sicily, steppe pastoralist ancestry arrived by ~2200 BC, in part from Iberia; Iranian-related ancestry arrived by the mid-second millennium BC, contemporary to its previously documented spread to the Aegean; and there was large-scale population replacement after the Bronze Age. In Sardinia, nearly all ancestry derived from the island's early farmers until the first millennium BC, with the exception of an outlier from the third millennium BC, who had primarily North African ancestry and who—along with an approximately contemporary Iberian—documents widespread Africa-to-Europe gene flow in the Chalcolithic. Major immigration into Sardinia began in the first millennium BC and, at present, no more than 56–62% of Sardinian ancestry is from its first farmers. This value is lower than previous estimates, highlighting that Sardinia, similar to every other region in Europe, has been a stage for major movement and mixtures of people.

After around 3000 BC, people with ancestry similar to the ancestry of those from the steppe north of the Black and Caspian Seas began to move westward, mixing with local Central European farmers and contributing up to three-quarters of the ancestry of peoples associated with the Corded Ware complex[1–3], and spreading in western Europe in association with the Bell Beaker complex[4,5]. In Iberia, steppe-pastoralist-related ancestry (hereafter, steppe ancestry) appeared in outlier individuals by around 2500 BC[5] and became fully mixed into the Iberian population by approximately 2000 BC[6]. On the island of Crete in the eastern Mediterranean, there was little if any steppe ancestry in any of the published individuals from the Middle to Late Bronze Age Minoan culture (dating to 2400–1700 BC), although these individuals derived about 15% of their ancestry from groups related to early Iranian herders (Iranian-related ancestry)[7]. However, steppe ancestry arrived in Crete and nearby Greece by the time of the Mycenaean culture (around 1600–1200 BC; Fig. 1).

In the islands of the central and western Mediterranean, the Bronze Age transition has not been investigated through the analysis of ancient DNA. The first permanent human presence in the Balearic Islands dates to ~2500–2300 BC[8,9]. Around 1200 BC, the Talaiotic culture was marked by intensified management of food resources and the appearance of monumental towers—the eponymous talaiots—some of which have suggestive similarities to the Sardinian nuraghi[10–12]. In turn, Bronze Age Nuragic Sardinian farmers also exchanged material goods with groups from the eastern Mediterranean[13]. Sardinia and Sicily were affected by the spread of the Beaker complex after ~2500 BC, while Sicily was affected by Aegean influences during the Mycenaean period[14–16]. The extent to which these cultural exchanges were accompanied by movements of people is an open question that we addressed by generating genome-wide ancient-DNA data from 61 individuals.

Results

The dataset. We prepared powder from petrous bones and teeth, extracted DNA[17–20] and converted it into double-stranded[21] or single-stranded libraries[22]. We treated all of the libraries with uracil-DNA

A full list of affiliations appears at the end of the paper.

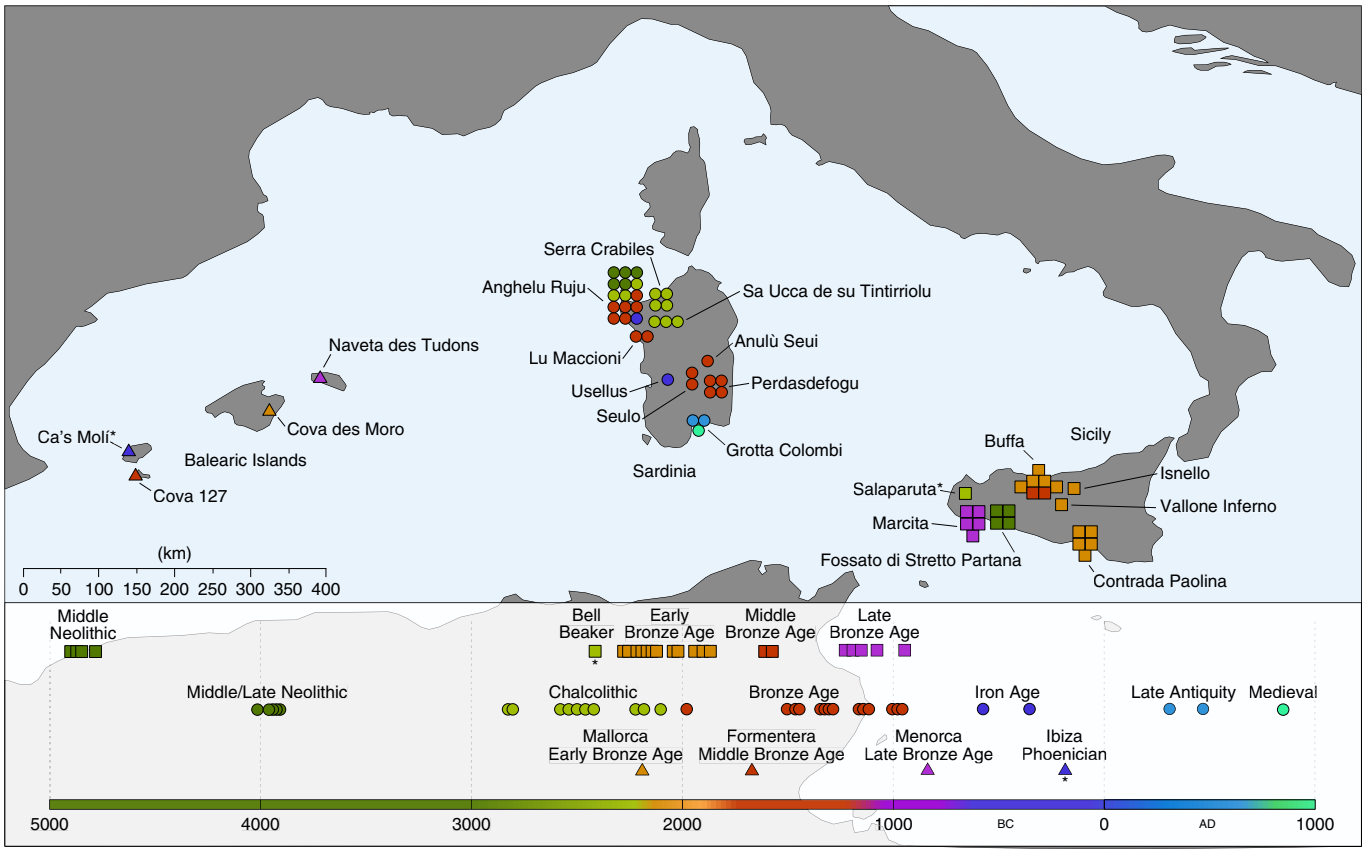

Fig. 1 | Geographical origins and temporal distribution of newly reported data. The 61 newly reported ancient individuals, along with a previously reported Beaker-associated Sicilian individual for whom we increase data quality and a reported Phoenician individual from Ibiza (both marked with an asterisk). Scale bar, 50 km sections.

glycosylase (UDG)[21] to greatly reduce the rate of cytosine-to-thymine errors, which are characteristic of ancient DNA. We enriched for sequences that overlapped approximately 1.24 million single-nucleotide polymorphisms (SNPs)[23,24], and then sequenced, thereby obtaining genome-wide data from 61 individuals from the Balearic Islands, Sardinia and Sicily, as well as increasing the data quality of a previously reported individual from Sicily[5] (Fig. 1, Supplementary Data 1, Supplementary Information). We assembled direct radiocarbon dates on bone samples from 46 individuals (Supplementary Data 2). We removed four individuals from the analysis for whom there was evidence of contamination on the basis of heterogeneity of the X chromosome (in males) or mitochondrial DNA, and one individual who was a first degree relative of another higher coverage individual[25]. For all analyses other than the by-sample analyses, we removed individuals with data for <100,000 SNPs. For the 49 individuals who remained for analysis after filtering, the median coverage at targeted SNPs on the autosomes was 2.91-fold (range, 0.11–12.13) and the median rate of cytosine-to-thymine damage in the terminal nucleotides was 11.7% (range, 3.7–25.6%)—within the range expected for authentic ancient DNA[21] (Supplementary Data 1). The genetic structure was similar when restricting to transversion SNPs that were not affected by characteristic ancient-DNA errors (Supplementary Fig. 1).

Overview of the genetic structure. We performed principal component analysis (PCA) on the newly reported ancient individuals merged with previously published data[1,2,4–7,26–62], projected onto the variation among 737 diverse present-day west Eurasians analysed at ~600,000 SNPs[7,40,63–65] (Fig. 2b, Supplementary Data 3). We also performed unsupervised clustering using ADMIXTURE[66] (Fig. 2a). We used qpWave[2] to evaluate whether each individual in turn was consistent with being from the same group as others from the same time period and region (that is, we tested whether they were consistent with forming a clade at $P < 0.01$), enabling us to create analysis groupings (Fig. 3, Supplementary Fig. 4). Where the analysis using qpWave was ambiguous, we performed refined tests to determine groupings (Supplementary Information).

In the Balearic Islands, the three Bronze Age individuals fall between the European Neolithic and Bronze Age clusters in the PCA; evidence of their steppe ancestry was also present in an ADMIXTURE component maximized in Yamnaya steppe pastoralists. qpWave revealed significant differences in ancestry between the Early Bronze Age (EBA) individual Mallorca_EBA and the Middle (MBA) and Late Bronze Age (LBA) individuals Formentera_MBA ($P = 0.001$) and Menorca_LBA ($P = 0.001$; Supplementary Table 1), respectively. Taking into account the partial evidence of genetic heterogeneity and the very different archaeological contexts of the three individuals, we treated each separately for analysis.

In Sicily, the four middle Neolithic (MN) individuals cluster with early European farmers and are genetically homogeneous and we therefore grouped them as Sicily_MN (Fig. 2, Supplementary Figs. 1 and 3). Relative to Sicily_MN, all of the Early Bronze Age Sicilians deviate in the PCA and ADMIXTURE analysis towards eastern groups with two Early Bronze Age outliers—Sicily_EBA11443 and Sicily_EBA8561, who clearly have steppe pastoralist admixture—and the remaining individuals dividing into a main Sicily_EBA grouping of four individuals and two subtly differentiated individuals, Sicily_EBA3123 ($P = 0.005$) and Sicily_EBA3124 ($P = 0.001$; Supplementary Tables 2 and 3). The two Middle Bronze Age individuals were consistent with being a clade and differentiated from

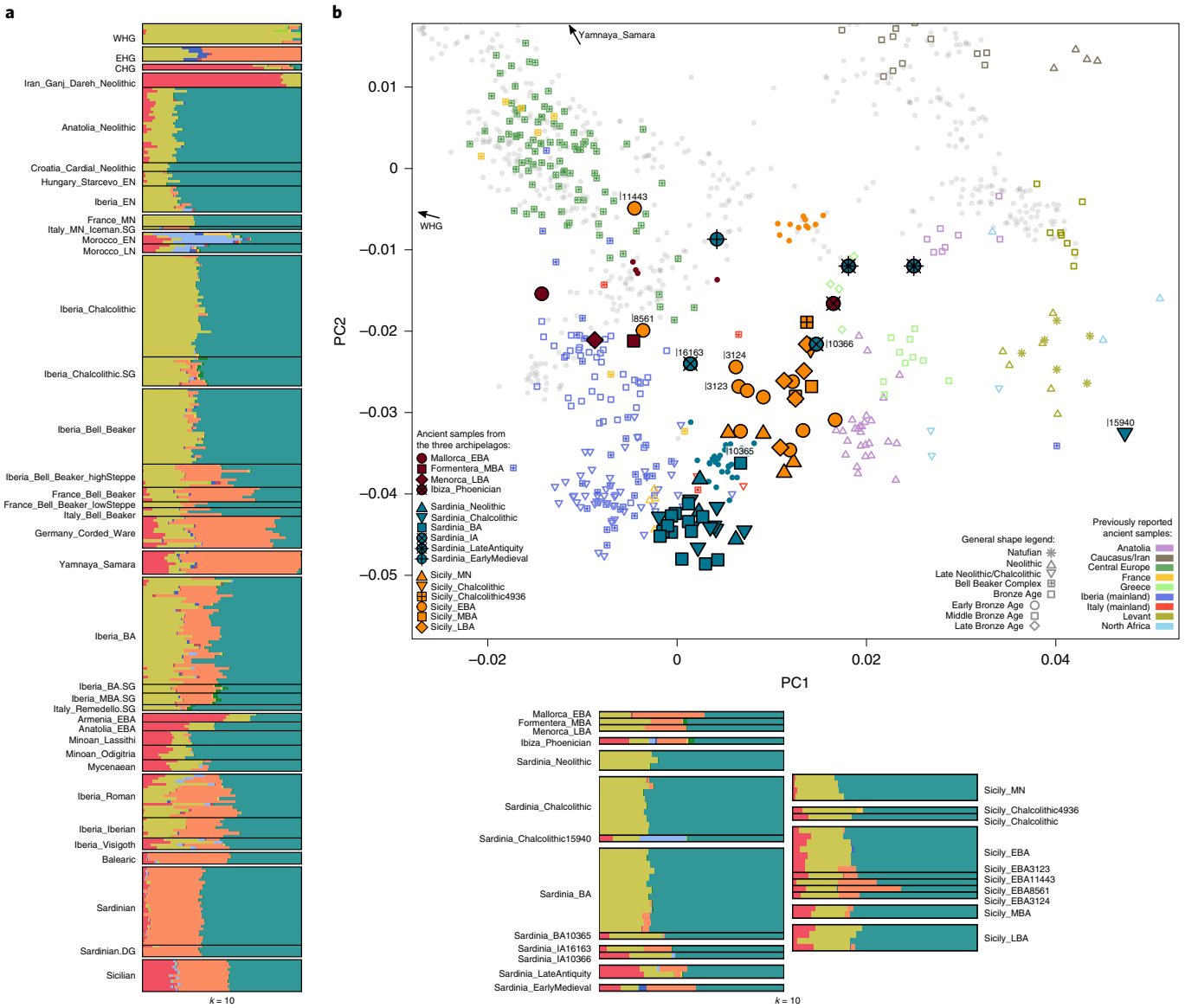

Fig. 2 | Overview of the genetic structure. a,**b**, The results for ancient Sardinians, Sicilians and Balearic islanders and other ancient and present-day populations according to unsupervised ADMIXTURE with $k = 10$ clusters (**a**) and a PCA with previously published data (non-filled symbols), projected onto the variation from present-day populations shown in solid-colour circles without outlines (**b**). Balearic islanders are indicated in maroon, Sicilians are indicated in orange, Sardinians are indicated in blue and all others are indicated in grey. Ibiza_Phoenician was published previously in ref. [59] and Sicily_Chalcolithic and Sicily_Chalcolithic4936 were published previously in ref. [5].

the Early Bronze Age individuals; we therefore grouped them as Sicily_MBA (Supplementary Table 4). All five Late Bronze Age individuals were consistent with being a clade at $P > 0.01$ (Sicily_LBA; Supplementary Table 5).

In Sardinia, we observed clustering with mainland European Early and Middle Neolithic farmers from the beginning of the Neolithic to the end of the Bronze Age; all but two individuals are indistinguishable in their ancestry components on the basis of analysis using qpWave (Fig. 2). The three Neolithic individuals were homogeneous and we therefore grouped them (Sardinia_Neolithic; Supplementary Table 6); all but one of the Chalcolithic individuals were homogeneous (Sardinia_Chalcolithic; Supplementary Tables 7 and 8), and all but one of the Nuragic Bronze Age individuals were homogeneous (Supplementary Tables 9 and 10). The two outliers were Sardinia_Chalcolithic15940 (radiocarbon dated to 2345–2146 calibrated (cal.) BC), who had significant affinity to Levantine and North African Neolithic individuals ($P < 10^{-12}$), and Sardinia_BA10365 (1643–1263 cal. BC), who had subtle evidence of eastern Mediterranean ancestry and was significantly differentiated from others of this period ($P = 0.000024$; Supplementary Tables 9 and 10). The two Iron Age individuals did not form a clade (Supplementary Table 11)—Sardinia_IA10366 (391–209 cal. BC) has evidence of Iranian-related ancestry, whereas Sardinia_IAI16163 (762–434 cal. BC) has evidence of steppe ancestry (Fig. 3). The two Sardinia_LateAntiquity individuals were consistent with forming a clade with each other and Sardinia_EarlyMedieval (Fig. 3, Supplementary Tables 12 and 13), but Sardinia_EarlyMedieval was separated from the earlier individuals in the PCA (Fig. 2) and dated to significantly later; we therefore analysed Sardinia_EarlyMedieval separately.

Balearic Islands. We used qpAdm[2,63] to decompose the ancestry of each analysis grouping into five distal sources that date to the

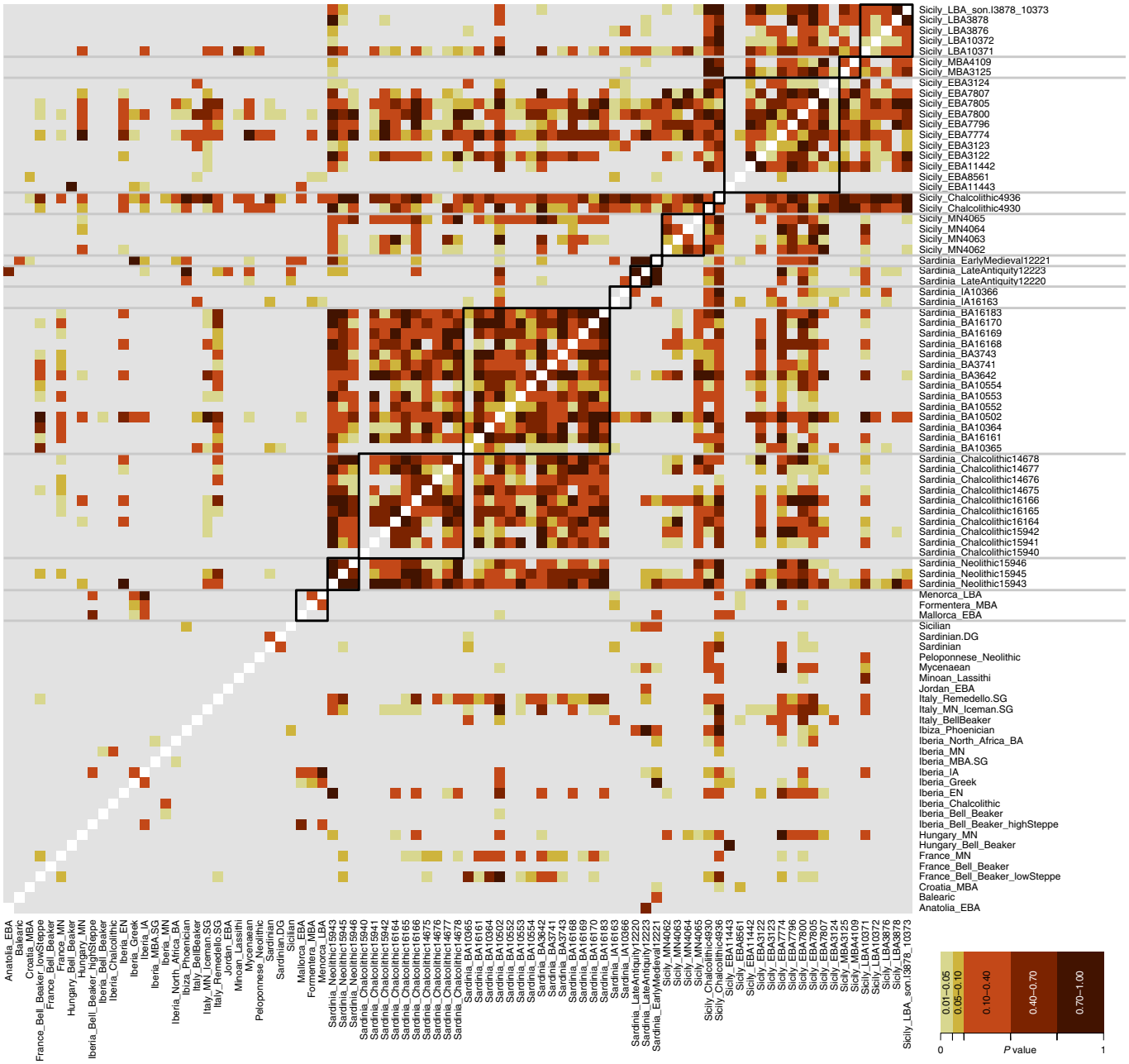

Fig. 3 | Pairwise qpWave testing to group individuals. Black lines represent groupings by location and period. Grey-coloured models have a *P* value of less than 0.01, and were rejected.

Late Neolithic (LN) or earlier: Anatolia_Neolithic, western hunter-gatherers (WHG), Iran_Ganj_Dareh_Neolithic, Yamnaya_Samara and Morocco_LN. We represented each of these sources with published ancient-DNA data. We caution that finding a fitting model in qpAdm should not be interpreted as an implication that the true source population came from the same location. Instead, a fit implies that the modelled population is consistent with being a mixture of groups that are derived from the same ancestral populations as the groups that we use as surrogates for them. We tested all of the possible two-, three-, four- and five-way mixture models, and—when more than one model of the same rank produced valid fits—we included the competing source among the set of outgroups. This 'model competition' approach[40,62] was designed to detect shared genetic drift between any test population and an outgroup that is not captured by one of the sources used in the model; this sometimes reveals that we are failing to correctly model a proximal ancestry source. We quote the most parsimonious model (as measured by the lowest number of ancestry sources) that fits at $P > 0.05$ and, when multiple models fit, we quote all of them (Fig. 4b, Supplementary Tables 14 and 16).

Mallorca_EBA dates to the earliest period of permanent occupation of the islands at around 2400 BC[10,67]. Using modelling, we estimated that $38.1 \pm 4.4\%$ of her ancestry derived from a source related to Yamnaya_Samara (Fig. 4, Supplementary Table 14). We next used qpAdm to identify proximal sources for the ancestry of Mallorca_EBA. Using our model competition approach, we tested a list of potential sources, motivated by closeness in space or time or showing cultural links, including all of them among the outgroups if they were not explicitly tested as sources. Mallorca_EBA can be modelled as a clade with the small subset of Iberian

Beaker-complex-associated individuals who carried steppe-derived ancestry[5] ($P=0.413$). We rejected models that did not involve an Iberian source, as we found no passing models when we used qpAdm to test non-Iberian sources for Mallorca_EBA with Iberians, including among the outgroups. Thus, the movements of people who brought steppe ancestry into Iberia overlapped with those who settled on the Balearic Islands. We caution that we have data from only a single individual from this period and her ancestry profile may not be representative of all of the early settlers. Future ancient-DNA sampling could reveal individuals with different proportions of steppe ancestry and without a specific Iberian connection.

Our estimates of steppe ancestry in the two later individuals from the Balearic Islands are lower—20.4 ± 3.4% for Formentera_MBA and 20.8 ± 3.6% for Menorca_LBA (Fig. 3, Supplementary Fig. 4, Supplementary Tables 1 and 14). These two individuals therefore represent mixtures of groups with a relatively large proportion of steppe ancestry, plausibly the population of which Mallorca_EBA was a part, and other groups with more early European farmer-related ancestry. This could be the result of early immigrants to the islands harbouring different proportions of steppe ancestry and then mixing, or later waves of gene flow from groups with relatively more European first-farmer-related ancestry. The proximal modelling of Formentera_MBA produced two-way fits with one source always bearing steppe ancestry (such as Mallorca_EBA, Iberia_Bell_Beaker_highSteppe or France_Bell_Beaker), and the other always bearing large proportions of Anatolian-Neolithic-related ancestry (fitting models include sources in mainland Italy, France or Sardinia; Supplementary Table 17). Menorca_LBA fits a one-way model with Formentera_MBA ($P=0.172$) and, therefore, may be directly continuous with that population (Supplementary Information).

To test for evidence of genetic links between peoples of the Talaiotic culture of the Balearic Islands and the Nuragic culture of Sardinia, we examined the fitting two-way admixture models for the Talaiotic-culture-associated individual Menorca_LBA. Although some of the models involved a Sardinian population as a fitting source, we also found fits for models that included France_Bell_Beaker_lowSteppe, Italy_Bell_Beaker, Italy_Remedello.SG or Italy_MN_Iceman.SG as sources, with Nuragic Sardinians included among the outgroups (Supplementary Table 18). Thus, our analysis does not produce any specific evidence of a genetic link between people of these two cultures. A caveat is that we have access to data from only a single Talaiotic-associated individual; it is also important to note that there can be cultural contact without movement of people.

Phoenician colonies were established in the Balearic Islands during the Iron Age. The Ibiza individual published previously[59] from a collective burial in a Punic hypogeum and dated to 361–178 cal. BC is not consistent with forming a clade with any of the Bronze Age Balearic individuals and has a qualitatively different ancestry profile; for example, a North African source of ancestry is required to obtain a fit (our model is 10.8 ± 2.7% Iran_Ganj_Dareh_Neolithic and 89.2 ± 2.7% Morocco_LN ancestry; Fig. 4, Supplementary Table 14). In proximal modelling, Ibiza_Phoenician also always requires Morocco_LN as one of the sources. Although some of these models include a Balearic Island Bronze Age source, it is possible that the Ibiza Phoenician individual has no ancestry at all from earlier Balearic peoples, as we fit her with models that have all of the Balearic Bronze Age individuals among the outgroups (for example, 17.0 ± 3.1% France_Bell_Beaker and 83.0 ± 3.1% Morocco_LN; Supplementary Table 19).

Modern Balearic individuals can be fit with only five distal sources, as a mixture of steppe, Iranian-related and North-African-related ancestry in addition to Anatolian-farmer-related and WHG. Thus, the present-day Balearic Islands reflect a mixture of disparate ancestry sources, some of which are likely to reflect movements in the past from the southern and eastern Mediterranean (Fig. 4, Supplementary Table 14). The uniparental haplogroups found in present-day Balearic populations include all of the haplogroups that have been found in the ancient individuals (mtDNA J2, H and U5, and Y chromosome R1b)[68,69], although none of these are unique to the ancient Balearic populations and, therefore, cannot be taken as a proof of local population continuity (Supplementary Data 1).

Sicily. In the Middle Neolithic, Sicilians harboured ancestry typical of early European farmers, which we fitted as a mixture of Anatolia_Neolithic and WHG (Figs. 2 and 4, Supplementary Table 14). We also approximately tripled coverage of a previously reported Beaker-complex-associated individual, and our reanalysis of these data confirmed the previous finding that there is no evidence of steppe ancestry[5].

In the Early Bronze Age, we found evidence of steppe ancestry by around 2200 BC. In distal qpAdm, the two outliers with the strongest evidence have 22.1 ± 3.6% steppe ancestry (Sicily_EBA8561) and 39.0 ± 3.5% steppe ancestry (Sicily_EBA11443); the latter individual is consistent with forming a clade with Mallorca_EBA ($P=0.245$), suggesting that they may harbour ancestry from a similar source (most plausibly Iberian; see below; Fig. 4a, Supplementary Table 14). For the main Early Bronze Age cluster of four individuals and two other outliers, we also fit steppe ancestry, albeit at the lower proportions of 14.1 ± 3.4% in Sicily_EBA3123, 13.5 ± 3.4% in Sicily_EBA3124 and 9.9 ± 2.2% in the main cluster of Sicily_EBA (Fig. 4, Supplementary Table 14, Supplementary Information). For Sicily_EBA and Sicily_EBA3123, we could not rule out an alternative model of Iranian-related ancestry rather than steppe as a third ancestry source, although we favour steppe models because of the results from the proximal modelling and the definitive presence of this ancestry in the two extreme outlier individuals. The presence of steppe ancestry in Early Bronze Age Sicily is also evident in Y-chromosome analysis, which revealed that three out of the five Early Bronze Age males carried haplogroup R1b1a1a2a1a2 (R1b-M269), which is associated with the first western Europeans who carried significant proportions of steppe ancestry (Supplementary Data 1). Two of these individuals carried the Y-chromosome haplogroup subtype R1b1a1a2a1a2a1 (Z195), which is now largely restricted to Iberia and has been hypothesized to have originated there during 2500–2000 BC[70]. A parsimonious scenario is that west-to-east gene flow from Iberia introduced these haplogroups into Sicily as well as to the Balearic Islands, where Y-chromosome haplogroup R1b-M269 is also found in Menorca_LBA.

We detected Iranian-related ancestry in Sicily by the Middle Bronze Age 1800–1500 BC, consistent with the directional shift of these individuals towards Minoans and Mycenaeans in the PCA (Fig. 2b); in distal modelling, Sicily_MBA requires 15.7 ± 2.6% of Iran_Ganj_Dareh_Neolithic-related ancestry ($P=0.060$; Fig. 4, Supplementary Table 14). Sources closer in time always require Minoan_Lassithi or Anatolia_EBA as a source (Supplementary Table 21). Modern southern Italians harbour Iranian-related ancestry[71], and our results show that this ancestry must have reached Sicily before the period of Greek political control when Sicily and southern Italy were part of Magna Graecia.

We modelled Sicily_LBA as 81.5 ± 1.6% Anatolia_Neolithic, 5.9 ± 1.6% WHG and 12.7 ± 2.1% Yamnaya_Samara (Fig. 4b, Supplementary Table 14). Although this distal modelling provides no hint of Iranian-related ancestry, modelling with sources closer in time supports Sicily_LBA having such ancestry through groups such as Anatolia_EBA or Minoan_Lassithi (Supplementary Table 22).

Our distal modelling of modern Sicilians requires not only that the two eastern ancestry sources that we have shown were present by the Bronze Age—10.0 ± 2.6% Yamnaya_Samara and 19.9 ± 1.4% Iran_Ganj_Dareh_Neolithic—but also a predominant component of North African ancestry (46.9 ± 5.6% Morocco_LN; Fig. 4,

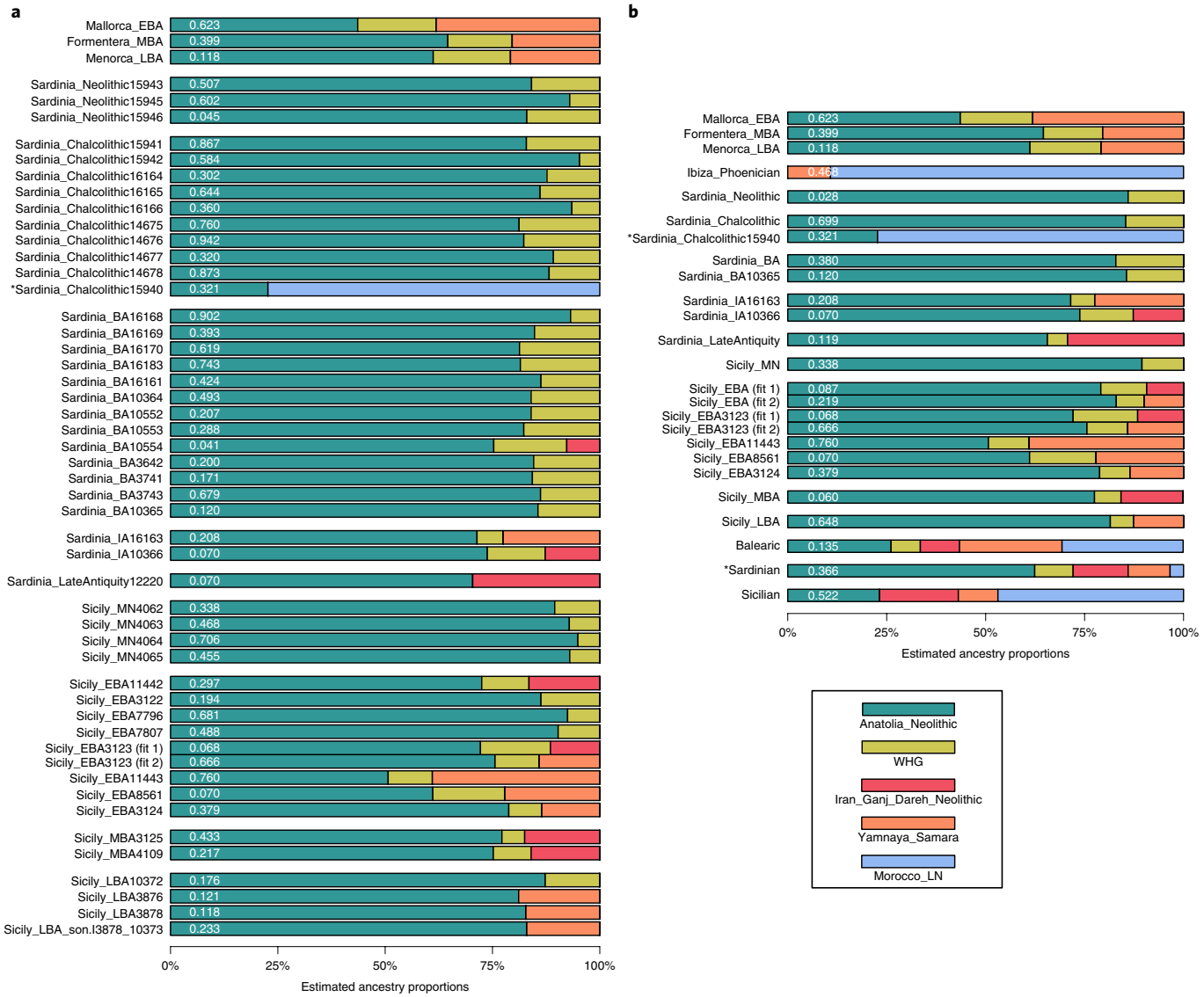

Fig. 4 | Distal modelling of ancestry proportions using qpAdm. a,**b**, Results are shown by individual (**a**) and analysis grouping (**b**) (the specific values are provided in Supplementary Tables 14 and 16); *P* values are shown within the bars. We show all of the valid models ($P > 0.05$) for the lowest fitting number of sources (some examples produced two valid models at $P > 0.05$). In **b**, we added a published individual (Ibiza_Phoenician) and modern groups (Balearic, Sardinian and Sicilian). The asterisks denote models using Morocco_EN instead of Morocco_LN to improve fits (Supplementary Information).

Supplementary Table 14). These results are consistent with most of the North-African-related ancestry having come into Sicily during the Iron Age and afterwards—a scenario that is further supported by our observation that modern Sicilians form a clade with Ibiza_Phoenician ($P = 0.060$) and the three most recent Sardinian individuals in our time series (Supplementary Information). Although these results are consistent in principle with a nearly complete ancestry turnover on the island since the Bronze Age, we cannot rule out the possibility that Bronze Age Sicilians made a more modest ancestry contribution to modern Sicilians. Uniparental markers in modern Sicilians overlap those from the Bronze Age, with Y-chromosome haplogroup R1b-M269 occurring at high frequency (~25%)[72] and mitochondrial haplogroups H, T, U and K also being present in the Bronze Age as well as in present-day groups[73].

Sardinia. In the qpAdm analysis, almost all of the Neolithic, Chalcolithic and Bronze Age individuals fit as descending from the same two deep ancestral sources (Anatolia_Neolithic and WHG) with similar and indistinguishable mixture proportion estimates of 82.9–86.0% (Fig. 4, Supplementary Table 14). Sardinia_Neolithic and Sardinia_Chalcolithic are consistent with being a clade ($P = 0.285$), as are Sardinia_Chalcolithic and Sardinia_BA ($P = 0.173$), further supporting a high degree of continuity (Supplementary Information). When we considered sources closer in time and space, Neolithic Sardinians modelled as a mixture of 74–77% France_MN, with the other portion related to Early Neolithic (EN) groups, with an estimate of 26% from Hungary_EN ($P = 0.051$) or 23% from Croatia_Cardial_Impressa_EN ($P = 0.051$; Supplementary Information). The true source of early Sardinian farmers is probably an unsampled group, potentially from a region such as northern Italy or Corsica for which there is little or no available ancient-DNA data. We detected no evidence of steppe ancestry, even in individuals buried in a Beaker context; this is similar to the pattern observed in Beaker-associated burials in Sicily and most in Iberia, a pattern that contrasts sharply with that in Central and Northern Europe, where Beaker-associated individuals had substantial steppe ancestry[5] (Supplementary Table 16, Supplementary Data 1).

Despite evidence of extraordinary ancestry continuity in Sardinia from the Neolithic through the Bronze Age, there are two notable outliers.

The most surprising is Sardinia_Chalcolithic15940 from the site of Anghelu Ruju, for whom we obtained a radiocarbon date of 2345–2146 cal. BC from the same bone sample that we analysed for DNA. We modelled this individual as 22.7 ± 2.4% Anatolia_Neolithic and 77.3 ± 2.4% Morocco_EN ($P = 0.321$). This individual is similar in ancestry composition to the approximately contemporary Iberian individual I4246 from the site of Camino de las Yeseras, radiocarbon dated to 2473–2030 cal. BC, who also had North-African-related ancestry as well as the same mtDNA haplogroup M1a1b1 and Y-chromosome haplogroup E1b1b1, which are both typical of North Africans[25] (Supplementary Table 14). The finding of African-to-European gene flow in both individuals shows that such movement was widespread across the Mediterranean long before the classical period when such gene flow became intensive and the ancestries had a larger demographic impact.

The second outlier is Sardinia_BA10365 (1643–1263 cal. BC; Fig. 2b), whom we fit as a mixture of a local Sardinian source and a second source carrying eastern Mediterranean-related (such as Mycenaean or Jordan_EBA) or steppe (Italy_Bell_Beaker or France_Bell_Beaker; Supplementary Table 23) ancestry. Uniparental marker analyses provided additional hints that the material culture exchange between Sardinians and the eastern Mediterranean was accompanied by some movement of people[74]. Sardinia_Chalcolithic15943 from Anghelu Ruju carried a rare U1a mitochondrial haplogroup, known from ancient individuals from the Early Bronze Age Balkans and western Asia[51,54,62] (Supplementary Data 1). Sardinia_BA10553 carried Y-chromosome haplogroup J2b2a (Supplementary Data 1), which today occurs at highest frequencies in the Balkans and the Middle East[75] and which was nearly unique to these regions during the Bronze Age and earlier[40,44,54,76].

The earliest definitive evidence of steppe and Iranian-related ancestries in Sardinia comes from two Iron Age individuals—Sardinia_IA16163 (762–434 cal. BC), with 22.5 ± 3.6% Yamnaya-related ancestry, and Sardinia_IA10366 (391–209 cal. BC), with 12.7 ± 3.5% Iran_Ganj_Dareh_Neolithic-related ancestry (Fig. 4, Supplementary Tables 14 and 24, Supplementary Information). The qpAdm models fit even when Bronze Age Sardinians are included in the outgroups, consistent with the hypothesis that these individuals, similar to the Phoenician from Ibiza in the Balearic Islands, had little ancestry from preceding local peoples. Iranian-related ancestry was even higher in at least some individuals in Late Antiquity, with the Sardinia_LateAntiquity cluster harbouring 29.3 ± 4.1% Iran_Ganj_Dareh_Neolithic-related ancestry ($P = 0.009$ for rejection of the alternative model that attempts to model ancestry as derived from the Yamnaya; Supplementary Table 14). Sardinia_LateAntiquity is consistent with being a clade with Ibiza_Phoenician ($P = 0.238$) as would be expected if this ancestry began to be introduced with the Phoenicians[77] (Supplementary Table 25). Although we do not model Sardinia_EarlyMedieval owing to his limited SNP coverage, his Y-chromosome haplogroup, E1b1b1b2, belongs to the same branch (E1b1b) as Sardinia_Chalcolithic15940 and Hellenistic-period Egyptians[48], consistent with a source in the eastern Mediterranean. Taken together, these results show that the five most recent Sardinians in our time series—from the Iron Age ($n = 2$), Late Antiquity ($n = 2$) and Early Medieval period ($n = 1$)—harboured minimal ancestry from Bronze Age, Chalcolithic or Neolithic Sardinians. All five individuals were from coastal sites, suggesting immigration from groups outside of Sardinia. Unsampled regions of Sardinia (possibly on the coast and almost certainly in the interior) probably retained high proportions of ancestry from pre-Iron-Age Sardinians in the Iron Age and later, as pre-Iron-Age Sardinians remain the single largest contributor to modern Sardinians (see below).

We were only able to model modern Sardinians by invoking four- and five-way distal admixture models, which include Iranian-related and North-African-related sources and are therefore substantially more complex than the three-way model for Sardinia proposed in the original paper that introduced the qpAdm method[2] (that study was not able to reject a three-way admixture model because it did not have access to our reference dataset of large numbers of ancient West Eurasians). With our more-refined modelling, the proportion of ancestry from the first farmers of Sardinia (comprising sources from Anatolian farmers and WHG)[2] is around 72.2% (62.5% Anatolian Neolithic + 9.7% WHG; Fig. 4, Supplementary Table 14). This value is much lower than that modelled by the first qpAdm study of ancestry proportions in Sardinians[2], which estimated that 87% of the island's ancestry could be derived from Neolithic farmers. It is also much less than the approximately 100% that was inferred in a study that found that modern Sardinians are a clade with ancient DNA from the late Neolithic Tyrolean Iceman[78]. Finally, it is also much less than the proportions of 90–100% that were estimated in a study of more than 3,500 modern Sardinians (in which the estimates vary across geographical regions)[79].

In fact, the true degree of population turnover was even higher. To test how much ancestry in modern Sardinians could be derived from local Sardinians from the Bronze Age and earlier, we examined the 27 modern Sardinians who were the source of many of these previous estimates[63]. These people are from the Gennargentu region and have some of the lowest amounts of steppe ancestry among modern Sardinians, who show variations in the proportions of different ancestries that reflect the geographical substructuring of modern Sardinians among different valleys and coastal and inland sites[79]. These modern Sardinians fit a four-way distal model with a local Sardinian ancestry source and three further sources—Iran_Ganj_Dareh_Neolithic, Yamnaya_Samara and Morocco_LN (Supplementary Table 26). We estimate that modern Sardinians retained between 56.3 ± 8.1% (when using Sardinia_Neolithic) and 62.2 ± 6.6% (with Sardinia_BA) of their ancestry from local populations, along with a substantial proportion of North African Morocco_LN-related ancestry that ranges between 22.7 ± 9.9% (when using Sardinia_Neolithic) and 17.1 ± 8.0% (when using Sardinia_BA). This North-African-related mixture is a plausible source for the sub-Saharan African admixture that has been detected in modern people from the island in multiple previous studies[79–82]. Even this range of estimates of ~56–62% is an upper bound, as we are forcing in pre-Iron-Age Sardinians as the only source of European farmer ancestry on the island in all of these models.

Taken together, these results confirm that the proportion of European first farmer-related ancestry is higher in Sardinia than elsewhere in Europe, while revealing a previously unknown major contribution from groups that arrived later. The ancestry from multiple sources is also evident in uniparental markers; for example, modern Sardinians carry haplogroups that were common on the island from the Neolithic through to the Bronze Age period (Y-chromosome haplogroup R1b1a(xR1b1a1a) and mtDNA haplogroups HV, JT and U[83]), as well as a high frequency of Y-chromosome haplogroup R1b-M269 that was absent in the Bronze Age and earlier[79,84]. A related ancient-DNA study[85] of Sardinia concurs that modern Sardinians harbour large proportions of ancestry from early Sardinian farmers as well as significant contributions from later periods that are larger than previously estimated. The study also agrees with our finding that many coastal Sardinian sites harboured people with little ancestry from earlier local Sardinians in the Iron Age and antiquity[85].

Discussion

We conclude by highlighting five observations. First, we identified Iberia as a key ancestry source for Bronze Age peoples of both the Balearic Islands and Sicily. Some early residents of the Balearic Islands probably derived at least part of their ancestry from

Iberia, as proximal models favour ancient Iberian sources. We also observed the characteristically Iberian Y-chromosome haplogroup R1b1a1a2a1a2a1 (Z195) in two Early Bronze Age Sicilians[70]. Thus, Iberia was not only a destination of east-to-west human movement, but was also an important source of west-to-east reflux[86]. However, the demographic history of Iberia during this period was also distinctive from the Mediterranean islands[25]; for example, no nearly complete replacement of paternal lineages occurred in Sicily, where both Neolithic and steppe-associated haplogroups persisted.

Second, our analysis shows that Iranian-related ancestry, which was widespread among the Aegean by the Middle Bronze Age in association with the Minoan and Mycenaean cultures, had also spread west into Sicily in a substantial proportion at least by the time of the Mycenaeans. One possibility is that this ancestry spread west along with the Mycenaean cultural expansion[87–90]. However, the presence of artifacts associated with the Sicilian Castellucian culture in Malta, Greece and Anatolia before the peak of the Mycenaean culture opens the possibility of earlier gene flows, a hypothesis that could be investigated once pre-Minoan data from the Aegean become available[87,91]. We found no evidence of substantial proportions of Iranian-related ancestry in the Balearic Islands or Sardinia before the Phoenician period, but this does not mean that these islands were isolated from the eastern Mediterranean; indeed, the opposite is true. Archaeological evidence shows that this was a period of unprecedented material culture exchange, with substantial westward flows of goods as reflected, for example, in the importation of Cypriot copper in Late Bronze Age Sardinia[92].

Third, our study highlights widespread human mobility from North Africa to Europe during the Chalcolithic and Bronze Age. Specifically, we identified an outlier individual in Sardinia with a large proportion of North-African-derived ancestry who dates to 2345–2146 cal. BC and has an ancestry profile that is similar to an approximately contemporary central Iberian individual dated to 2473–2030 cal. BC[25] (and a Bronze Age individual from Iberia dated to 1932–1697 cal. BC, who carried North-African-related ancestry in admixed form[25]). Together, 1.6% of the 191 individuals from Mediterranean Europe in our analysis dataset from between 5,000–3,000 years ago have evidence of ancestry from North African migrants in the few generations before they lived[61].

Fourth, our analysis documents the impact of the Phoenician and Greek colonial periods as well as subsequent immigration on the islands of the western Mediterranean[93]. For example, all six individuals in our analysis dataset from the post-Bronze Age period are consistent with having no ancestry from earlier local groups. In conjunction with previous findings, the emerging picture is that, from the Iron Age onward, the coastal regions of the western Mediterranean were characterized by ethnically segregated populations of immigrants and local groups who co-existed in geographical proximity; indeed, there is direct documentation of this from the Greek colony of Empúries in northeast Iberia, where two clusters of genetically distinct individuals co-existed, consistent with the historical descriptions by Strabo[25]. In some regions, such as the Balearic Islands and Sicily, our data are consistent with a nearly complete replacement of the pre-Iron-Age populations (although we cannot rule out a degree of local continuity for either set of islands).

Finally, our co-analysis of modern and ancient Sardinians questions the common view that Sardinians are well described as isolated descendants of Europe's first farmers[78]. Our ancient-DNA time transect does reveals that Neolithic, Chalcolithic and Bronze Age Sardinians had a typical early European farmer ancestry profile that persisted longer on the island than anywhere else in Europe studied to date and, in this sense, our study supports previous findings that Sardinia is special in Europe in retaining more of its first-farmer ancestry. However, immigrants from the eastern Mediterranean and North Africa made a substantial demographic impact in Sardinia from the Iron Age onwards, just as they did in other parts of the coastal western Mediterranean. Therefore, all five post-Bronze Age individuals from Sardinia in our time series (all from coastal sites) have no evidence of pre-Bronze-Age Sardinian ancestry, and they clearly mixed with the previously established populations to contribute at least ~38–44% of the ancestry of modern Sardinians (with North-African-derived ancestry estimated at ~17–23%). Thus, rather than being fully sheltered from admixture and migration since the Neolithic, Sardinia—similar to almost all other regions in Europe—has been a major stage for movement and mixture of peoples.

Methods

Laboratory work and bioinformatics analysis. We extracted powder from skeletal samples at dedicated ancient-DNA facilities at the University of Vienna, University College Dublin, the University of Florence, the University of Palermo and Harvard Medical School[17,19,20,94,95] (Supplementary Data 1). We treated all but one DNA extract (S4420.E1.L1) with UDG to remove characteristic ancient-DNA damage and thereby greatly reducing the rate of damage-induced errors[96]. For all but one sample, we performed DNA extraction at Harvard Medical School, sometimes using silica-coated magnetic beads to support robotic clean-ups (instead of silica column clean-ups that were used for manual DNA extraction)[17,19,20]. We converted these DNA extracts to individually double-barcoded (or double-indexed) libraries using either double-strand ligation[21] or single-strand ligation[22], in most cases assisted by a robotic liquid handler (Supplementary Data 1). For one of the samples, we performed DNA extraction[17,18] and double-indexed library preparation using double-stranded ligation in Florence[21].

We initially screened some libraries by enriching them for the human mitochondrial genome[97] and about 3,000 nuclear SNPs (Supplementary Data 1, mtDNA+3000SNPs) using synthesized baits (CustomArray). We performed sequencing using an Illumina NextSeq500 instrument with 2×76 cycles and read the indices with 2×7 cycles; for some of the libraries, we performed sequencing using a HiSeq X Ten with 2×101 cycles and read the indices with 2×8 cycles. We assigned sequences on the basis of the library-specific barcodes/indices. We used one out of two bioinformatics pipelines to process the data, as specified in Supplementary Data 1. In pipeline 1, we merged read pairs that overlapped by at least 15 bp, allowing for up to one mismatch using SeqPrep (https://github.com/jstjohn/SeqPrep; representing each overlapping base by the higher quality base); whereas, for pipeline 2, we allowed one mismatch when the forward and reverse base had a quality of ≥20, or 3 mismatches when quality was <20 (always retaining the base with higher quality but, when mismatches were found, base quality was defined as the difference in base qualities). We used either SeqPrep (pipeline 1) or a custom script (https://github.com/DReichLab/ADNA-Tools; pipeline 2) to computationally trim adapters and barcodes. We mapped the merged sequences to the reconstructed human mtDNA consensus sequence[98] using bwa (for pipeline 1, v.0.6.1; for pipeline 2, v.0.7.15-r1140)[99] with the parameters --n 0.01 and --l 16500 (pipeline 1) or -n 0.01, -o 2 and -l 16500 (pipeline 2). We removed duplicate sequences that had the same orientation and same start and stop positions using a custom script (pipeline 1) or the Broad Institute's Picard MarkDuplicates tool (http://broadinstitute.github.io/picard/) (pipeline 2). We restricted to sequences of at least 30 bp with a mapping quality of ≥30 and to bases with base quality of ≥30. We assessed the data for authenticity by computing the damage rate at the terminal cytosines[21] and by estimating the rate of mismatches to the consensus mitochondrial sequence using contamMix[23] (v.1.0–10 for pipeline 1 and v.1.0–12 for pipeline 2). We determined mitochondrial haplogroups using HaploGrep2 (ref. [100]; Supplementary Data 4).

To evaluate whether differences between the two bioinformatics pipelines were introducing artefacts into our analysis—for example, by causing sequences processed using pipeline 1 to match each other at a higher rate than those sequenced by pipeline 2—we performed two sets of tests. For the first set of tests, we separated the Sardinia_BA individuals by processing pipeline (3 from pipeline 1 and 10 from pipeline 2). We computed symmetry f_4-statistics of the form f_4(Mbuti.DG, Test, Sardinia_BA_pipeline1, Sardinia_BA_pipeline2) using 12 groups including modern Sardinian, Mycenaean, Iberia_BA.SG, Menorca_LBA as 'Test' (Supplementary Data 5). None of the tests produced a significant deviation from zero (the largest had a $|Z| = 1.126$ deviation from zero, showing that, at least for Sardinia_BA, the processing of samples with different pipelines did not affect the results.

For our second set of tests, we analysed data for 20 previously published individuals[25,62] whose sequencing data we processed using both pipelines. To test for artifactual attraction of samples processed by pipeline 1 and samples processed by pipeline 2, we computed 'ABBA' and 'BABA' counts and associated P values for the differences in the rates of these counts for the set of samples (individual1_pipeline1, individual1_pipeline2; individual2_pipeline1, individual2_pipeline2). For example, a BABA count corresponds to a case in which the sequences match for the two individuals processed using pipeline 1, and mismatch the sequences at the same position for the two individuals processed using pipeline 2 (a BABA count is the direction expected for a bias, and we therefore expect to observe an

excess of BABA over ABBA counts if there is a bias). We computed ABBA and BABA counts for all 190 possible comparisons, and assessed statistical significance on the basis of a one-sided binomial test for an excess of BABA over ABBA counts (assuming conservatively that the counts are independent). Only one test produced a P value of less than 0.05 ($P=0.022$), which was not significant after Bonferroni correction for the 89 pairwise tests in which we had sufficient coverage for the pair of samples to have power in principle to detect a signal at the $P<0.05$ level (that is, ABBA + BABA was at least 5). The results are provided in Supplementary Data 6.

We enriched the screened libraries with promising quality for 1,233,013 targeted SNPs ('1240k' target capture)[2,24] and newer libraries for the mitochondrial genome together with the 1240k targets ('1240k+' target capture). We sequenced and processed the data as described for the mtDNA, except that we mapped to the human reference genome hg19. We estimated contamination on the basis of the ratio of Y- to X-chromosome sequences (confirming that all of the individuals in the dataset had a ratio that was consistent with a male or a female) as well as the rate of heterozygosity at X chromosome positions (which is only valid as an estimate of contamination in males, who should have no X chromosome variation[101]). For pipeline 1, we determined each SNP in each individual by a randomly sampled sequence covering the nucleotide with mapping quality of ≥30 and base quality of ≥30; whereas, for pipeline 2, we used mapping quality of ≥10 and base quality of ≥20. We restricted analysis to individuals with a coverage of at least 20,000 SNPs (when analysing populations represented by a single person, we required a coverage of at least 100,000 SNPs).

Radiocarbon dating and quality control. We obtained 43 accelerator mass spectrometry (AMS) radiocarbon dates (^{14}C) at the Pennsylvania State University (PSU) Radiocarbon Laboratory, as well as an additional three direct dates from other laboratories (the dates were from 48 distinct samples as we dated two samples twice). Here we provide a description of sample processing at PSU, as it is the source of most of our dates (the sample description is adapted from the standard description of this processing; for the other samples, we refer readers to the published protocols). As a precaution at PSU, we removed possible contaminants (conservants and adhesives) by sonicating all of the bone samples in successive washes of ACS grade methanol, acetone and dichloromethane for 30 min each at room temperature, followed by three washes in Nanopure water to rinse. We extracted bone collagen and purified using a modified Longin method with ultrafiltration (>30 kDa gelatin[102]). If collagen yields were low and amino acids were poorly preserved, we used a modified XAD process (XAD Amino Acids[103]). For quality assurance, we measured carbon and nitrogen concentrations and C/N ratios of all of the extracted and purified collagen/amino acid samples using a Costech elemental analyzer (ECS 4010). We evaluated sample quality by percentage of crude gelatin yield, percentage of C, percentage of N and C/N ratios before AMS ^{14}C dating. C/N ratios for all directly radiocarbon samples fell between 3.1 and 3.3, indicating excellent preservation[104]. We combusted collagen/amino acid samples (~2.1 mg) for 3 h at 900 °C in vacuum-sealed quartz tubes with CuO and Ag wires. Sample CO_2 was reduced to graphite at 550 °C using H_2 and a Fe catalyst, and we drew off reaction water with $Mg(ClO_4)_2$ (ref. [105]). We pressed graphite samples into targets in aluminium boats and loaded them onto a target wheel and performed all ^{14}C measurements using a modified National Electronics Corporation compact spectrometer with a 0.5 MV accelerator (NEC 1.5SDH-1). We corrected the ^{14}C ages for mass-dependent fractionation with measured δ^{13}C values[106] and compared with samples of Pleistocene whale bone (backgrounds, >48,000 ^{14}C BP), late Holocene bison bone (~1,850 ^{14}C BP), late AD 1800s cow bone and OX-2 oxalic acid standards. We calibrated ^{14}C ages using OxCal v.4.3 (ref. [107]) and the IntCal13 northern hemisphere curve[108]. The stable carbon and nitrogen isotope measurements that we obtained do not indicate a large marine dietary component in these individuals despite the fact that they come from island populations and we therefore did not perform a correction of the dates for the marine reservoir effect.

Uniparental haplogroup determination. We determined mitochondrial haplogroups using HaploGrep2 (ref. [100]) and phylotree[109] (build 17) on the data from the mitochondrial enrichment experiments[105] (Supplementary Data 4). We restricted sequences and base qualities to values of ≥30, and built a consensus mitochondrial genome sequence with samtools and bcftools[110] (v.0.1.19-96b5f2294a) using a majority rule and a minimum coverage of 1 and trimming 2 bp from the end of each sequence. We further restricted the data for each sample to the damaged reads as determined by pmdtools (using a minimum pmdscore of 3) and repeated the calling. In almost every case in which there was sufficient post-damage restricted coverage to give a confident haplogroup call, the calls for the damage-restricted data matched the calls obtained when we did not perform damage-restriction. For Y-chromosome haplogroup assessment, we used only base qualities of ≥30, identifying the most derived mutations using the nomenclature of the International Society of Genetic Genealogy (http://www.isogg.org) v.11.110.

Dataset assembly. Our basic dataset included 3,359 individuals of whom 2,182 were modern[7,40,63–65] and 1,177 were ancient individuals from previous publications[1,2,4–7,26–62]; we combined these individuals with the newly reported 57 samples that passed screening (Supplementary Data 3). We performed all subsequent analysis on autosomal data.

PCA. We used a subset of 591 modern and 1,266 ancient west Eurasians for PCA using smartpca from the EIGENSOFT package[111](v.7.2.1) using the default parameters with two exceptions. We used the option lsqproject: YES to project all of the ancient individuals onto the eigenvectors computed from modern vectors. The approach of projecting each ancient sample onto patterns of variation learned from modern samples is useful for ancient-DNA analysis, as it means that we can use data from a large fraction of SNPs covered in each individual and therefore maximize the information about ancestry that would be lost in approaches that require restriction to a potentially smaller number of SNPs for which there is intersecting data across low coverage ancient individuals[112]. We used the option shrinkmode: YES to remap the points for the samples used to generate the PCA onto the positions where they would be expected to fall if they had been projected (thereby allowing the projected and non-projected samples to be appropriately covisualized). We used a dataset containing only transversions to assess the robustness of our qualitative inferences to bias due to errors induced by damaged ancient DNA (Supplementary Fig. 1).

Population structure analysis. We ran ADMIXTURE[66] (v.1.2.3) after pruning to remove 1 SNP each in pairs of SNPs in linkage disequilibrium using PLINK1.9 (ref. [113]) and the option --indep-pairwise 200 25 0.4, leaving 315,235 SNPs. We ran ADMIXTURE from $K=5$ to $K=15$, with 10 random-seeded replicates for each value of K. We used cross-validation by adding the option --cv to find the runs with the lowest errors (Supplementary Fig. 2). For each K value, we retained the replicate with the lowest error. We present results for $K=10$, as we empirically found that this is the value of K with lowest cross-validation error that also showed clear distinctions between ancient western, eastern and Caucasus hunter-gatherer backgrounds, while maximizing the Neolithic-Anatolian-associated component. We also performed ADMIXTURE, restricting to transversion SNPs, and obtained qualitatively similar results, suggesting that ancient-DNA damage probably did not strongly bias our findings (Supplementary Fig. 1).

f_4-statistics. We used ADMIXTOOLS[63] (v.4.1) to compute f_4 symmetry statistics (qpDstat). We used Mbuti.DG as the outgroup, and computed statistics of the form f_4(Mbuti.DG, X; Y, Z), where X is our test population (or individual) and Y and Z are pairs of populations (or individuals) that we tested for sharing alleles at an equal rate with X. We used the options f4mode: YES and printsd: YES.

qpWave and qpAdm. We used qpWave and qpAdm from ADMIXTOOLS[63] (v.4.1) to estimate admixture coefficients and to model our individuals/populations as mixtures of groups that we propose as clades with the true source population. We used a base outgroup set including the following individuals/populations: Mbuti.DG, Ust_Ishim, CHG, EHG, ElMiron, Vestonice16, MA1, Israel_Natufian and Jordan_PPNB. Extra populations were included in each test to improve accuracy when using populations with similar ancestries (a detailed description is provided in the Supplementary Information). When visualizing the results, we present the results with the fewest sources that also have the highest probability (although we also report alternative models with the same number of sources that fit at $P>0.05$). We used the option allsnps: YES to specify that the f_4-statistics, which are the basis of qpWave and qpAdm analyses, should be computed using the union of SNPs with coverage in all four groups that contribute to each f_4-statistic. We set a minimum threshold of 100,000 SNPs when modelling single individuals.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All raw data are available at the European Nucleotide Archive under accession number PRJEB35980 and at https://reich.hms.harvard.edu/datasets.

Code availability

All custom code used in this study is provided at https://github.com/DReichLab/ADNA-Tools.

Received: 21 March 2019; Accepted: 8 January 2020;
Published online: 24 February 2020

References

1. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
2. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
3. Kristiansen, K. Re-theorising mobility and the formation of culture and language among the corded ware culture in Europe. *Antiquity* **91**, 334–347 (2017).
4. Cassidy, L. M. et al. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl Acad. Sci. USA* **113**, 368–373 (2016).

5. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
6. Martiniano, R. et al. The population genomics of archaeological transition in west Iberia: investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet.* **13**, e1006852 (2017).
7. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
8. Alcover, J. A. The first Mallorcans: prehistoric colonization in the western Mediterranean. *J. World Prehist.* **21**, 19–84 (2008).
9. Ramis, D. in *The Cambridge Prehistory of the Bronze and Iron Age Mediterranean* (eds Knapp, A. B. & van Dommelen, P.) 40–56 (Cambridge Univ. Press, 2014).
10. Ramis, D. Animal exploitation in the early prehistory of the Balearic Islands. *J. Isl. Coast. Archaeol.* **13**, 269–282 (2018).
11. Plantalamor, L. & Van Strydonck, M. *La Cronologia de la Prehistòria de Menorca: (Noves datacions de 14C). Treballs del Museu de Menorca* Vol. 20 (Museu de Menorca, 1997).
12. Lull, V., Mico, R., Rihuete, C. I. & Risch, R. Los cambios sociales en las Islas Baleares a lo largo del II milenio. *Cypsela* **15**, 123–148 (2004).
13. Holt, E. in *Forging Identities. The Mobility of Culture in Bronze Age Europe* Vol. 1 (eds Suchowska-Ducke, P., Reiter, S. S. & Vandkilde, H.) 193–202 (British Archaeological Reports, 2015).
14. Ugas, G. *L'alba Dei Nuraghi* (Fabula, 2005).
15. Sestieri, A. M. B. in *The Oxford Handbook of European Bronze Age* (eds Harding, A. & Fokkens, H.) 653–667 (Oxford Univ. Press, 2013).
16. Sarno, S. et al. Ancient and recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the Mediterranean. *Sci. Rep.* **7**, 1984 (2017).
17. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
18. Damgaard, P. B. et al. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015).
19. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
20. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
21. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Proc. R. Soc. B* **370**, 20130624 (2015).
22. Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).
23. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
24. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
25. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
26. Keller, A. et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 698 (2012).
27. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
28. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
29. Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
30. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
31. Skoglund, P. et al. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
32. Günther, T. et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl Acad. Sci. USA* **112**, 11917–11922 (2015).
33. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
34. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
35. Olalde, I. et al. A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol. Biol. Evol.* **32**, 3132–3142 (2015).
36. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
37. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
38. Hofmanová, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
39. Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
40. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
41. Martiniano, R. et al. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun.* **7**, 10326 (2016).
42. Schiffels, S. et al. Iron age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016).
43. González-Fortes, G. et al. Paleogenomic evidence for multi-generational mixing between Neolithic farmers and Mesolithic hunter-gatherers in the Lower Danube basin. *Curr. Biol.* **27**, 1801–1810 (2017).
44. Haber, M. et al. Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am. J. Hum. Genet.* **101**, 274–282 (2017).
45. Jones, E. R. et al. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576–582 (2017).
46. Lipson, M. et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
47. Saag, L. et al. Extensive farming in Estonia started through a sex-biased migration from the steppe. *Curr. Biol.* **27**, 2185–2193 (2017).
48. Schuenemann, V. J. et al. Ancient Egyptian mummy genomes suggest an increase of sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* **8**, 15694 (2017).
49. Unterländer, M. et al. Ancestry and demography and descendants of Iron Age nomads of the Eurasian steppe. *Nat. Commun.* **8**, 14615 (2017).
50. Amorim, C. E. G. et al. Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* **9**, 3547 (2018).
51. Damgaard, P. et al. The first horse herders and the impact of Early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
52. Fernandes, D. M. et al. A genomic Neolithic time transect of hunter-farmer admixture in central Poland. *Sci. Rep.* **8**, 14879 (2018).
53. Fregel, R. et al. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc. Natl Acad. Sci. USA* **115**, 6774–6779 (2018).
54. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
55. Mittnik, A. et al. The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
56. Valdiosera, C. et al. Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc. Natl Acad. Sci. USA* **115**, 3428–3433 (2018).
57. van de Loosdrecht, M. et al. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
58. Veeramah, K. R. et al. Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc. Natl Acad. Sci. USA* **115**, 3494–3499 (2018).
59. Zalloua, P. et al. Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. *Sci. Rep.* **8**, 17567 (2018).
60. Feldman, M. et al. Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat. Commun.* **10**, 1218 (2019).
61. González-Fortes, G. et al. A western route of prehistoric human migration from Africa into the Iberian Peninsula. *Proc. R. Soc. B* **286**, 20182288 (2019).
62. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
63. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
64. Pickrell, J. K. et al. The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).
65. Qin, P. & Stoneking, M. Denisovan ancestry in east Eurasian and native American populations. *Mol. Biol. Evol.* **32**, 2665–2674 (2015).
66. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
67. Ramis, D., Alcover, J. A., Coll, J. & Trias, M. The chronology of the first settlement of the Balearic Islands. *J. Mediterr. Archaeol.* **15**, 3–24 (2002).
68. Picornell, A., Gómez-Barbeito, L., Tomàs, C., Castro, J. A. & Ramon, M. M. Mitochondrial DNA HVRI variation in Balearic populations. *Am. J. Phys. Anthropol.* **128**, 119–130 (2005).
69. Adams, S. M. et al. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am. J. Hum. Genet.* **83**, 725–736 (2008).
70. Solé-Morata, N. et al. Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. *Sci. Rep.* **7**, 7341 (2017).
71. Raveane, A. et al. Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in southern Europe. *Sci. Adv.* **5**, eaaw3492 (2019).

72. Di Gaetano, C. et al. Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur. J. Hum. Genet.* **17**, 91–99 (2009).
73. Sarno, S. et al. An ancient Mediterranean melting pot: investigating the uniparental genetic structure and population history of Sicily and southern Italy. *PLoS ONE* **9**, e96074 (2014).
74. Holt, E. M. *Economy and Environment in Complex Societies: A Case Study from Bronze Age Sardinia* (Univ. Michigan, 2013).
75. Magoon, G. R. et al. Generation of high-resolution a priori Y-chromosome phylogenies using 'next-generation' sequencing data. Preprint at *bioRxiv* https://doi.org/10.1101/000802 (2013).
76. Wang, C.-C. et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
77. Matisoo-Smith, E. et al. Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: a story of settlement, integration, and female mobility. *PLoS ONE* **13**, e0190169 (2018).
78. Sikora, M. et al. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS Genet.* **10**, e1004353 (2014).
79. Chiang, C. W. K. et al. Genomic history of the Sardinian population. *Nat. Genet.* **50**, 1426–1434 (2018).
80. Moorjani, P. et al. The history of African gene flow into southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373 (2011).
81. Loh, P.-R. et al. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
82. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
83. Olivieri, A. et al. Mitogenome diversity in Sardinians: a genetic window onto an Island's past. *Mol. Biol. Evol.* **34**, 1230–1239 (2017).
84. Morelli, L. et al. A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with cultural rather than demic diffusion of agriculture. *PLoS ONE* **5**, e10419 (2010).
85. Marcus, J. H. et al. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. *Nat. Commun.* https://doi.org/10.1038/s41467-020-14523-6 (2020).
86. Sangmeister, E. Die datierung des rickstroms der Glockenbecker und ihre auswirkung auf die chronologie der Kupferzeit in Portugal. *Palaeohistoria* **12**, 395–407 (1966).
87. Holloway, R. *The Archaeology of Ancient Sicily* (Routledge, 2000).
88. D'Agata, A. L. Interactions between Aegean groups and local communities in Sicily in the Bronze Age: the evidence from pottery. *Stud. Micenei ed Egeo-Anatolici* **42**, 61–83 (2000).
89. Shelton, K. in *The Oxford Handbook of the Bronze Age Aegean* (ed. Kline, E.) 139–148 (Oxford Univ. Press, 2012).
90. Alberti, G. Issues in the absolute chronology of the Early-Middle Bronze Age transition in Sicily and southern Italy: a Bayesian radiocarbon view. *J. Quat. Sci.* **28**, 630–640 (2013).
91. Heyd, V. in *The Oxford Handbook of the European Bronze Age* (ed. Harding, A.) 47–67 (Oxford Univ. Press, 2013).
92. Sabatini, S. Late Bronze Age oxhide and oxhide-like ingots from areas other than the Mediterranean: problems and challenges. *Oxf. J. Archaeol.* **35**, 29–45 (2016).
93. Aubet, M. E. & Turton, M. *The Phoenicians and the West: Politics, Colonies and Trade* (Cambridge Univ. Press, 1997).
94. Pinhasi, R. et al. Optimal ancient DNA Yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
95. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* **14**, 1194–1205 (2019).
96. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
97. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).
98. Behar, D. M. et al. A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
99. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
100. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
101. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
102. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
103. Lohse, J. C., Madsen, D. B., Culleton, B. J. & Kennett, D. J. Isotope paleoecology of episodic mid-to-late Holocene bison population expansions in the southern Plains, U.S.A. *Quat. Sci. Rev.* **102**, 14–26 (2014).
104. van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999).
105. Santos, G. M., Southon, J. R., Druffel-Rodriguez, K. C., Griffin, S. & Mazon, M. Magnesium perchlorate as an alternative water trap in AMS graphite sample preparation: a report on sample preparation at Kccams at the University of California, Irvine. *Radiocarbon* **46**, 165–173 (2004).
106. Stuiver, M. & Polach, H. A. Discussion reporting of 14C data. *Radiocarbon* **19**, 355–363 (1977).
107. Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720–730 (2013).
108. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
109. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
110. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
111. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
112. Skoglund, P. et al. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).
113. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

Acknowledgements

This manuscript is dedicated to the memory of S. Tusa of the Soprintendenza del Mare in Palermo, who would have been an author of this study had he not tragically died in the crash of Ethiopia Airlines flight 302 on 10 March 2019. We thank Z. Zhang for database support; the Soprintendenza BBCCAA Palermo and R. Schicchi (director of Museum of Castelbuono) for facilitating access to important skeletal materials. D.M.F. was supported by an Irish Research Council grant GOIPG/2013/36. Radiocarbon work was supported in part by the NSF Archaeometry program BCS-1460369 (to D.J.K. and B.J.C). C.L.-F. was supported by Obra Social La Caixa and by FEDER-MINECO (BFU2015-64699-P and PGC2018-095931-B-100). D.C. was supported by grant 20177PJ9XF MIUR PRIN 2017. D.Reich is an Investigator of the Howard Hughes Medical Institute, and his ancient-DNA laboratory work was supported by National Science Foundation HOMINID grant BCS-1032255, a National Institutes of Health grant GM100233, an Allen Discovery Center grant, and grant no. 61220 from the John Templeton Foundation.

Author contributions

D.M.F., D.Reich and R.P. conceived the study. D.M.F., E.C., C.C., G.C., M.C., V.F., M.Lozano, E.M., M.Michel, R.M.M., D.Ramis, M.R.P., V.S., P.S., L.T., M.T.-N., C.L.-F., L.S., D.C., A.C., M.Lucci, G.G., F.C., G.S. and R.P. excavated, assembled and/or studied the osteological material. D.M.F., O.C., N.R., N.B., M.F., B.G., M.Lari, M.Micheletti, A.Modi, M.N., F.C., J.O., K.A.S., K.S., K.M., C.S., K.T.Ö. and S.V. performed laboratory work under the supervision of N.R., D.C. and R.P. J.C. provided computing resources. B.J.C. performed radiocarbon analysis under the supervision of D.J.K. D.M.F., I.O., R.B., S.M. and M.Mah performed bioinformatics and population genetics analysis with input from A.Mittnik, I.L., N.P. and D.Reich.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-1102-0.

Correspondence and requests for materials should be addressed to D.M.F., D.C., R.P. or D.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria. [2]Earth Institute and School of Archaeology, University College Dublin, Dublin, Ireland. [3]CIAS, Department of Life Sciences, University of Coimbra, Coimbra, Portugal. [4]Department of Genetics, Harvard Medical School, Boston, MA, USA. [5]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. [6]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [7]Area 52 Research Group, School of Biology and Environmental Science/Earth Institute, University College Dublin, Dublin, Ireland. [8]Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA, USA. [9]Institut Català de Paleoecologia Humana i Evolució Social (IPHES), Tarragona, Spain. [10]Àrea de Prehistòria, Universitat Rovira i Virgili (URV), Tarragona, Spain. [11]Dipartimento di Biologia, Università di Firenze, Florence, Italy. [12]Centre for Applied Bioanthropology, Institute for Anthropological Research, Zagreb, Croatia. [13]DANTE Laboratory of Diet and Ancient Technology, Sapienza University of Rome, Rome, Italy. [14]Superintendency of Archaeology, Fine Arts and Landscape for the provinces of Sassari and Nuoro, Sassari, Italy. [15]Superintendency of Archaeology, Fine Arts and Landscape for the city of Cagliari and the provinces of Oristano and South Sardinia, Cagliari, Italy. [16]Instituto Internacional de Investigaciones Prehistóricas de Cantabria, Universidad de Cantabria-Gobierno de Cantabria-Banco Santander, Santander, Spain. [17]Dipartimento di Scienze della Vita e dell' Ambiente, Sezione di Neuroscienze e Antropologia, Universita' di Cagliari, Cagliari, Italy. [18]Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche, Università di Palermo, Palermo, Italy. [19]Dipartimento Culture e Società, Università di Palermo, Palermo, Italy. [20]Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Torino, Italy. [21]CNR-Istituto di Geologia Ambientale e Geoingegneria c/o Dipartimento di Scienze della Terra, Sapienza Università di Roma, Rome, Italy. [22]Independent researcher, Palma de Mallorca, Spain. [23]Museo Archeologico Regionale Antonino Salinas, Palermo, Italy. [24]Instituto de Ciencias del Patrimonio (Incipit-CSIC), Santiago de Compostela, Spain. [25]McDonald Institute for Archaeological Research and Homerton College, University of Cambridge, Cambridge, UK. [26]Department of Anthropology, Natural History Museum Vienna, Vienna, Austria. [27]Department of Anthropology, University of California, Santa Barbara, Santa Barbara, CA, USA. [28]Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona, Spain. [29]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [30]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy. [31]Max Planck–Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Cambridge, MA, USA. [32]Present address: Department of Anthropology, University of California, Santa Cruz, Santa Cruz, CA, USA. [33]Present address: Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [34]Present address: Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA. [35]Present address: Department of Genetics, Harvard Medical School, Boston, MA, USA. *e-mail: daniel.fernandes@univie.ac.at; david.caramelli@unifi.it; ron.pinhasi@univie.ac.at; reich@genetics.med.harvard.edu

nature research

Corresponding author(s): David Reich

Last updated by author(s): 22-12-2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	N/A
Data analysis	cutadapt v1.5 BWA v.0.6.1 BWA v.0.7.15-r1140 samtools v0.1.19-96b5f2294a mapDamage v2.0.8 ADMIXTOOLS 4.1 EIGENSOFT v.7.2.1 ADMIXTURE v.1.23 OxCal v.4.3 SeqPrep PicardTools 2 contamMix v.1.0-10 contamMix v.1.0-12 pmdtools PLINK 1.9

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw data are available at the European Nucleotide Archive and the National Center for Biotechnology Information under the accession number [to be included upon paper acceptance] and at https://reich.hms.harvard.edu/datasets.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	62 (sample size depended on sample availability and DNA sequencing success)
Data exclusions	Some samples were excluded from the population genetic analysis due to contamination or low coverage. This is described in more detail in the main manuscript.
Replication	N/A
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging